Machine Learning 1

Lecture 1.2 - What is Machine Learning?

Erik Bekkers

(Bishop 1.0 and 1.1)

Slide credits: Rianne van den Berg

Image credit: Kirillm | Getty Images

What is machine learning?

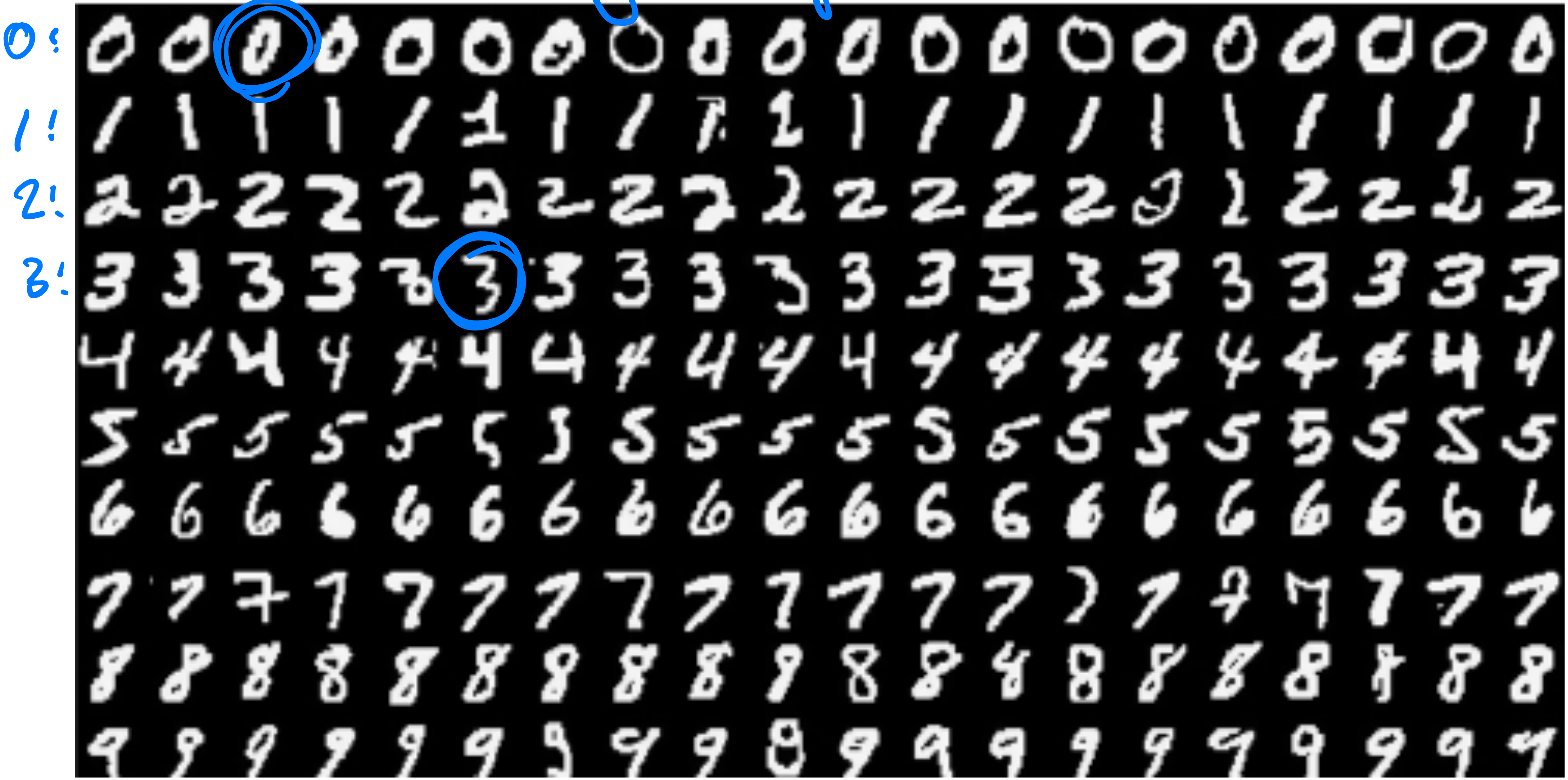
“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.”

- Tom M. Mitchell

Machine Learning, Tom Mitchell, McGraw Hill, 1997

E: Experience

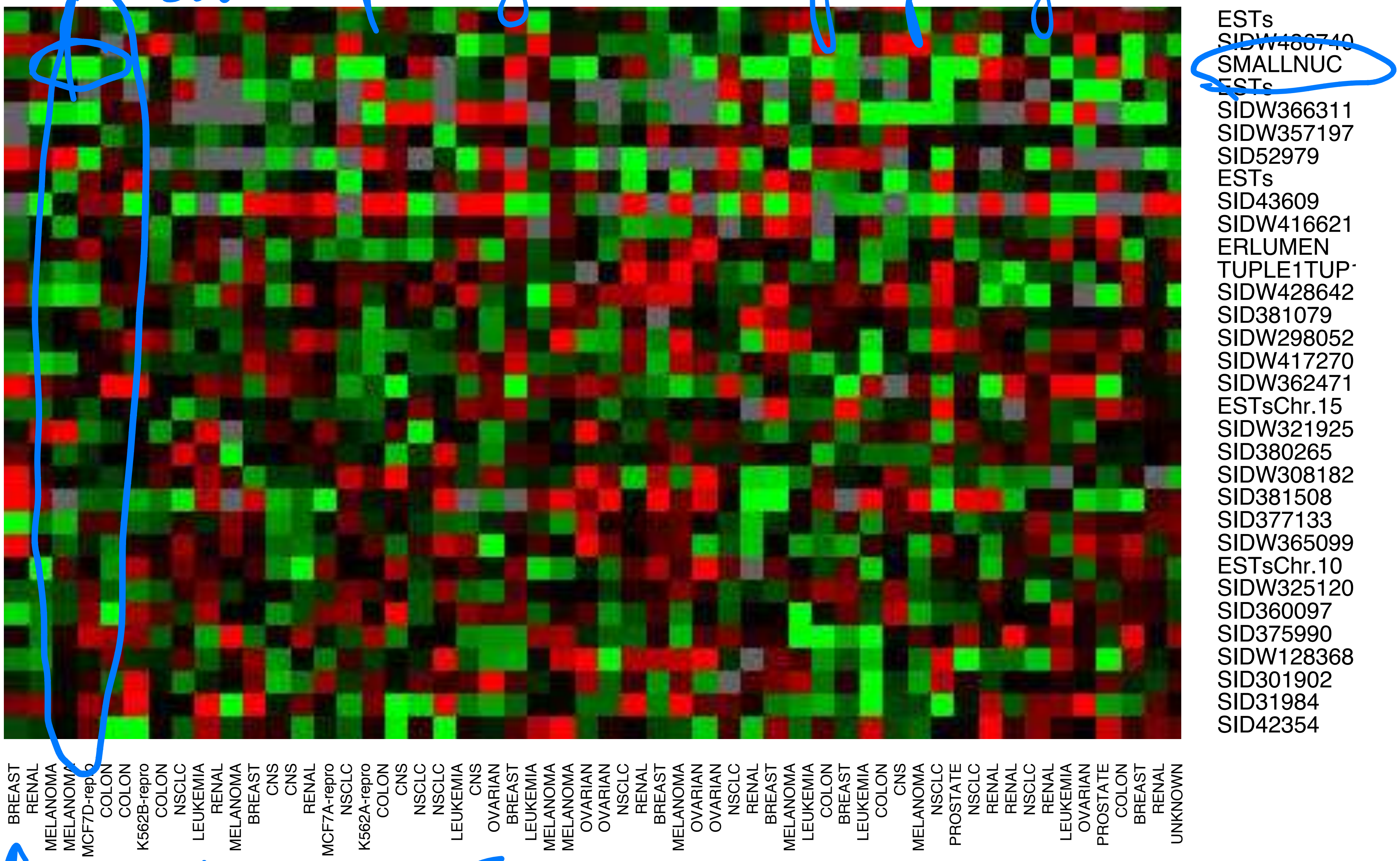
Handwritten digits of size 28x28



MNIST dataset

E: Experience

heatmap for activity of a gene

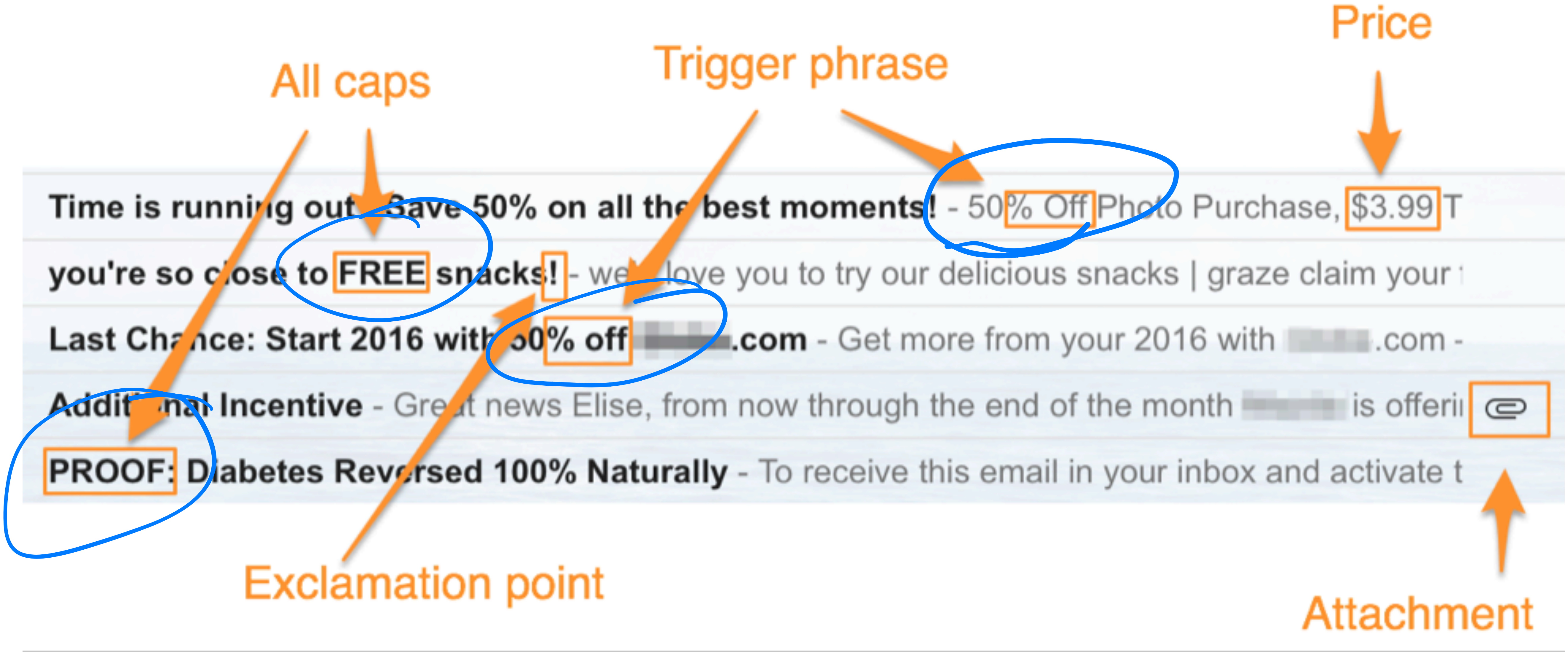


gene

Tumor

Expression matrix of genes (rows) for 64 human tumor samples (columns). [source: ESL 1.3]

E: Experience



Time is running out. Save 50% on all the best moments! - 50% Off Photo Purchase, $3.99 T

you're so close to FREE snacks! - we'd love you to try our delicious snacks | graze claim your

Last Chance: Start 2016 with 50% off [redacted].com - Get more from your 2016 with [redacted].com -

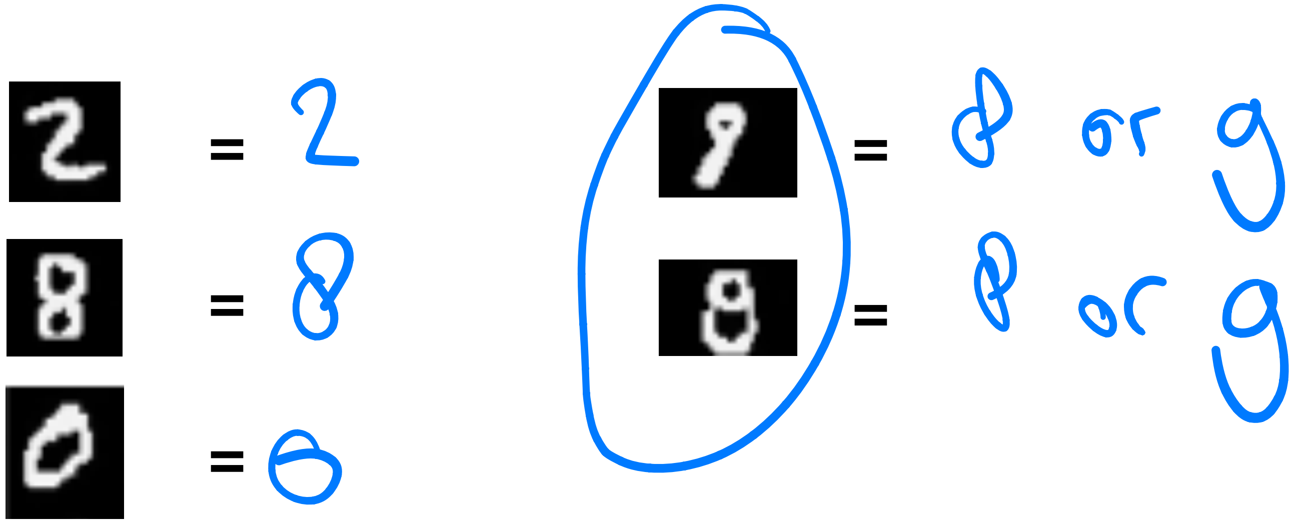
Additional Incentive - Great news Elise, from now through the end of the month [redacted] is offeri

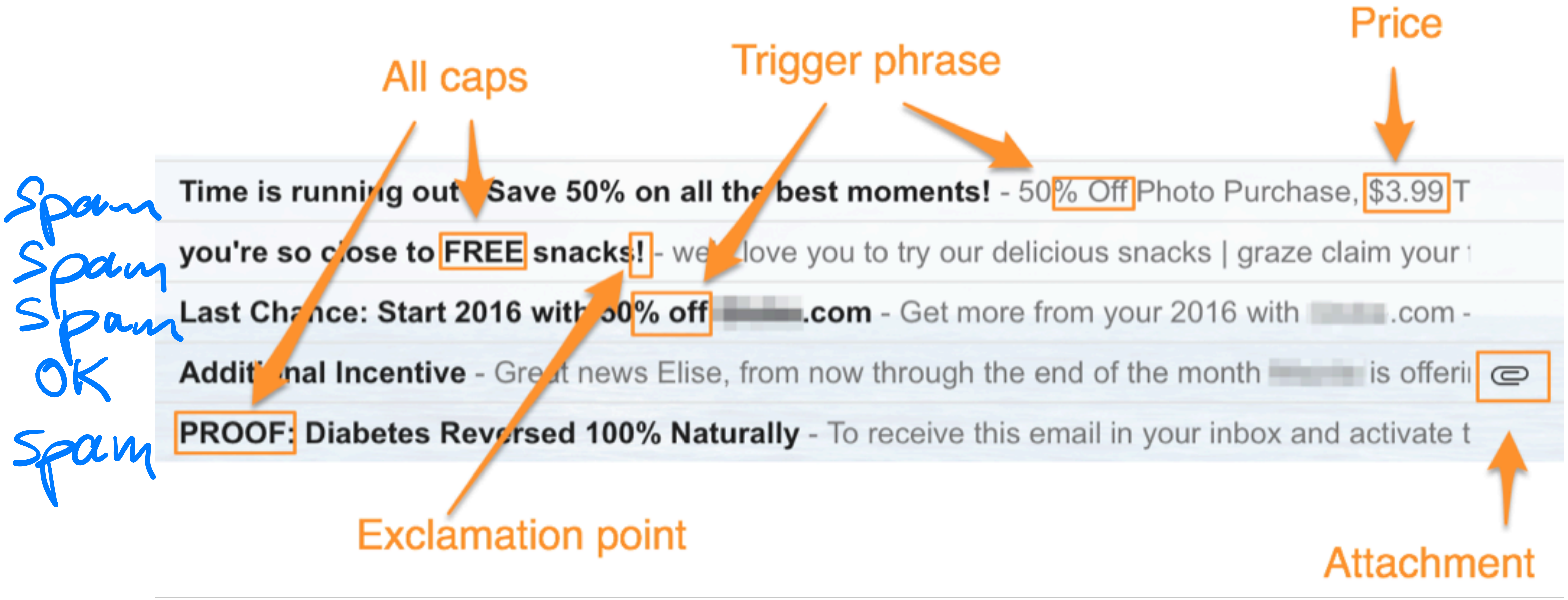
PROOF: Diabetes Reversed 100% Naturally - To receive this email in your inbox and activate t

Examples of spam emails. [source: Yesware]

T: Class of tasks

Classification:




Price

All caps

Trigger phrase

Spam Time is running out! Save 50% on all the best moments! - 50% Off Photo Purchase, $3.99 T

Spam you're so close to FREE snacks! - we'd love you to try our delicious snacks | graze claim your

Spam Last Chance: Start 2016 with 50% off [redacted].com - Get more from your 2016 with [redacted].com -

OK Additional Incentive - Great news Elise, from now through the end of the month [redacted] is offeri

Spam PROOF: Diabetes Reversed 100% Naturally - To receive this email in your inbox and activate t

Exclamation point

Attachment

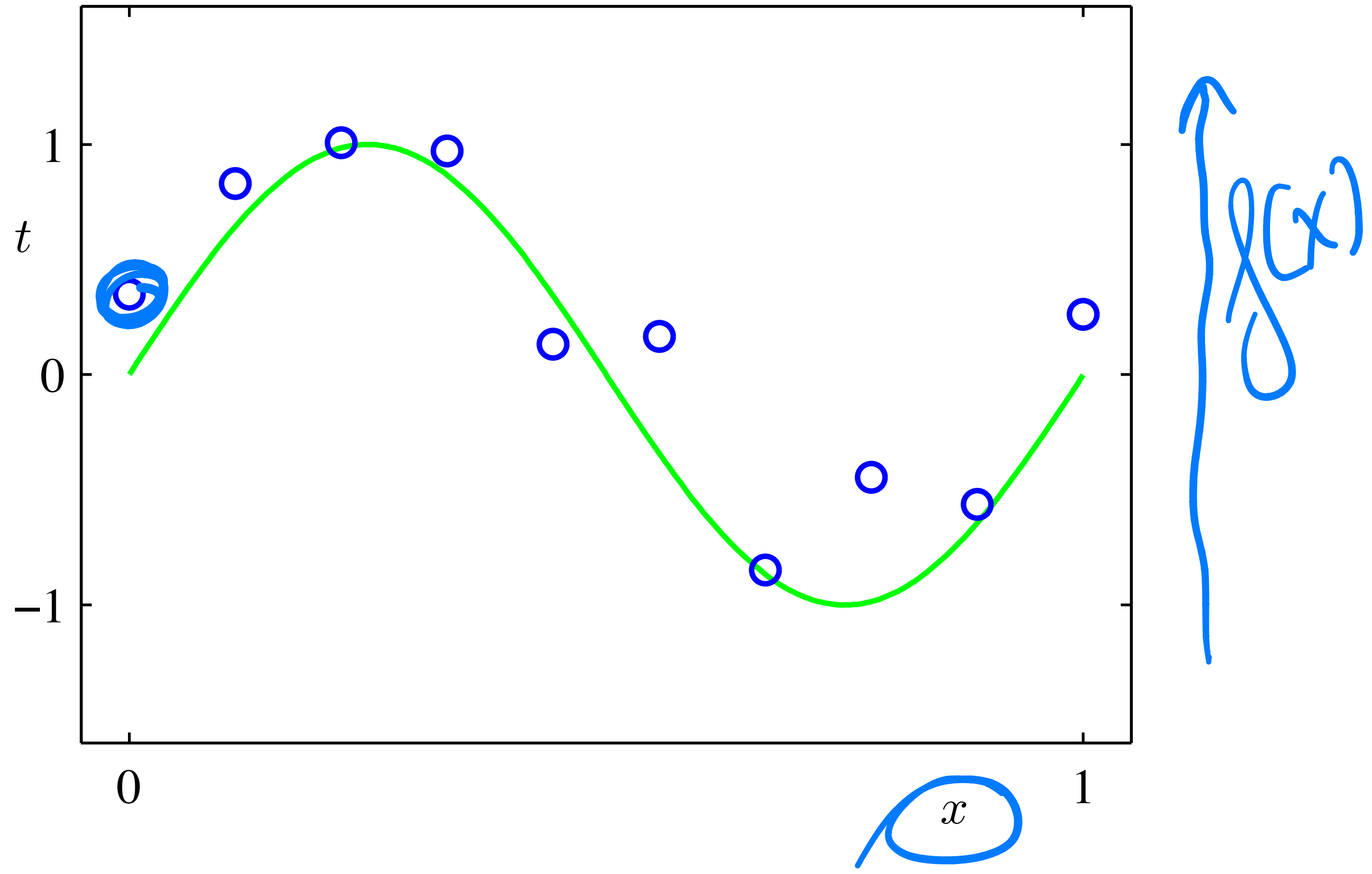
T: Class of tasks

Regression

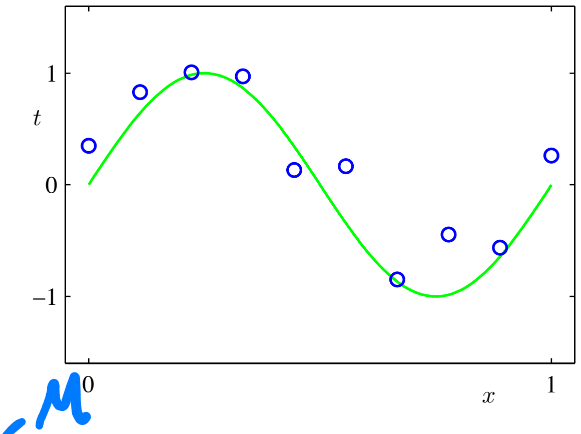
input : x

target : $t = \sin(2\pi x) + \varepsilon$

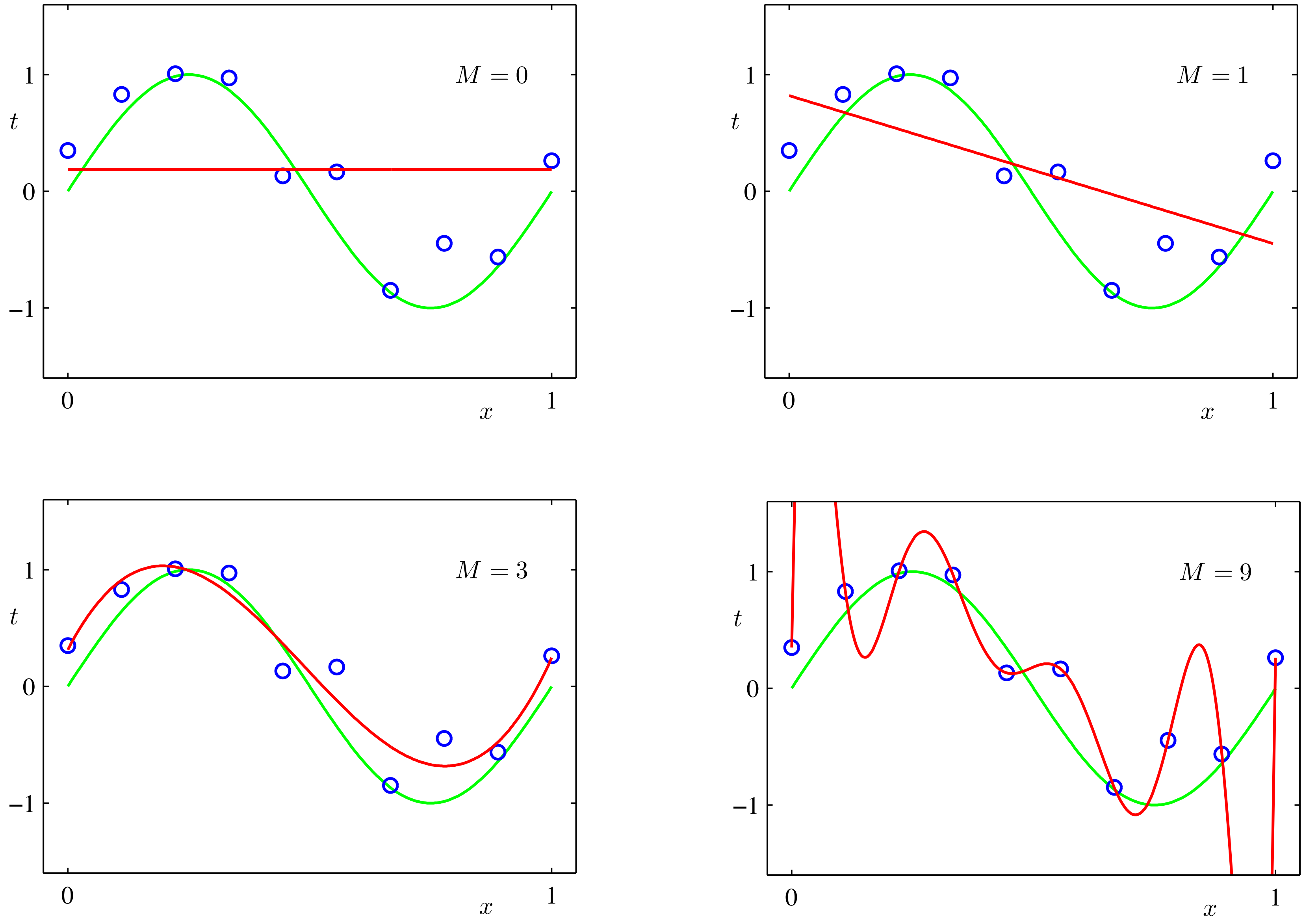
noise : $\varepsilon \sim N(0, 1)$



T: Class of tasks



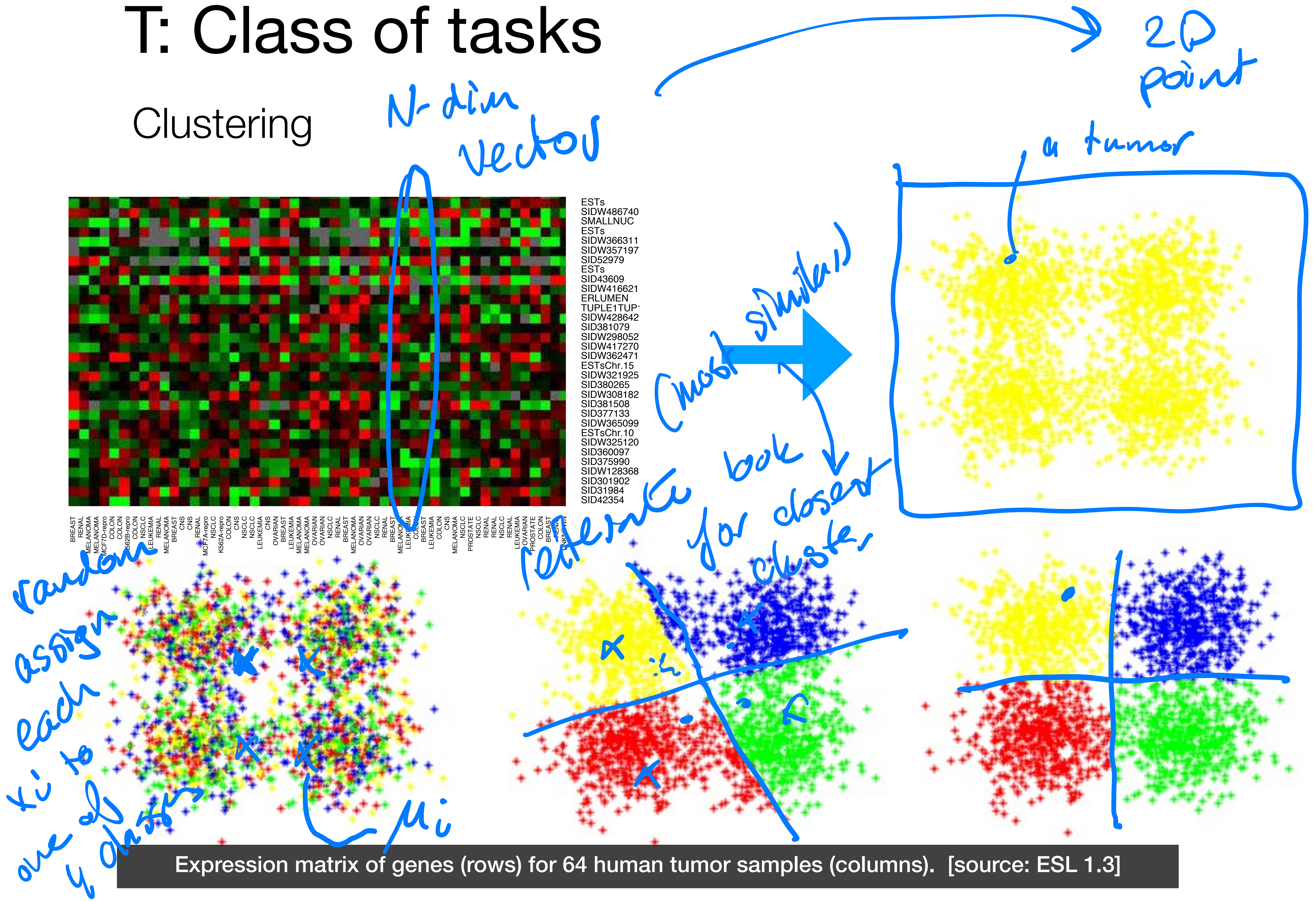
Regression $f(x) = w_0 + w_1 x + w_2 x^2 + \dots + w_M x^M$



Polynomials of order *M* (red) fit to data constructed as t = sin(2πx) + ε (green)

T: Class of tasks

Clustering



N-dim vector

2D point

a tumor

(most similar)

reference book for closest cluster

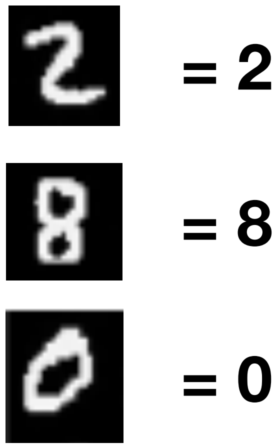
random assign each x_i to one of 4 classes

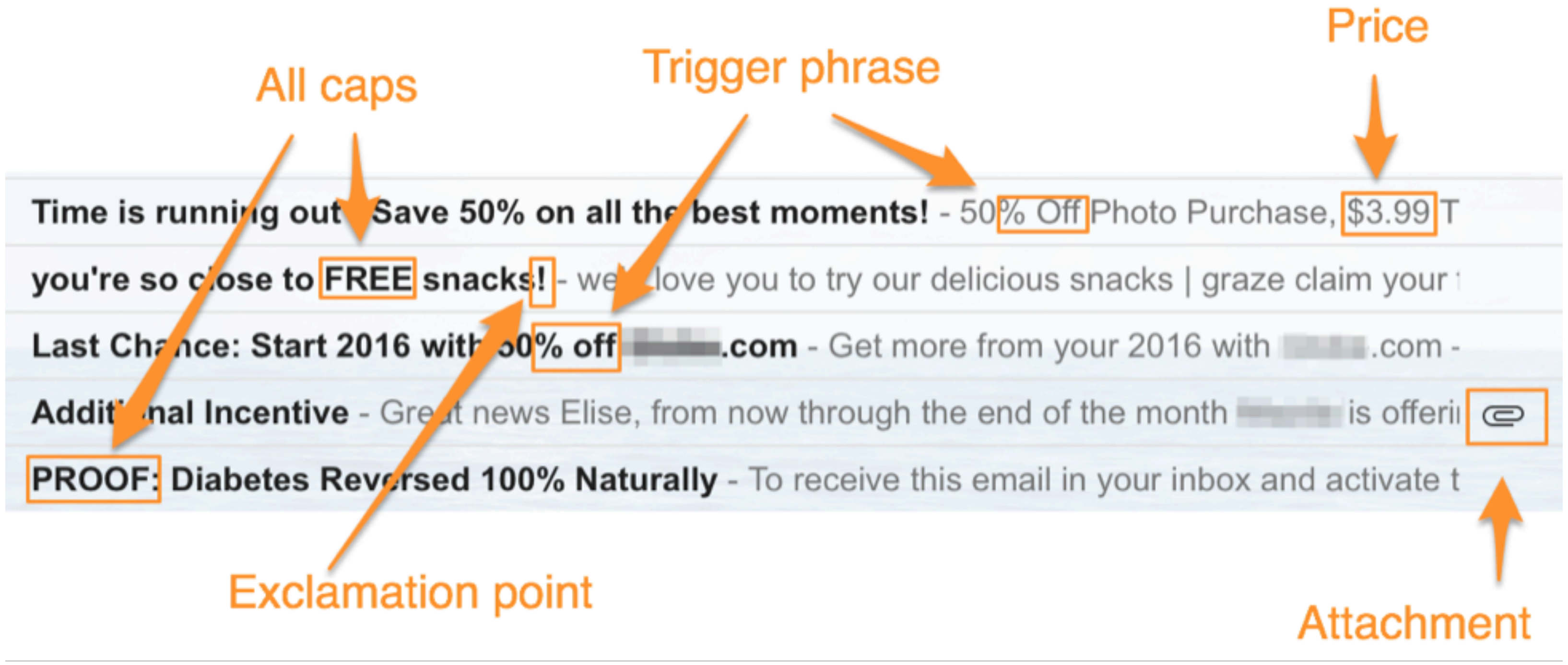
μ_i

Expression matrix of genes (rows) for 64 human tumor samples (columns). [source: ESL 1.3]

P: Performance measure

Classification





Time is running out! Save 50% on all the best moments! - 50% Off Photo Purchase, $3.99 T

you're so close to FREE snacks! - we love you to try our delicious snacks | graze claim your

Last Chance: Start 2016 with 50% off [redacted].com - Get more from your 2016 with [redacted].com -

Additional Incentive - Great news Elise, from now through the end of the month [redacted] is offeri

PROOF: Diabetes Reversed 100% Naturally - To receive this email in your inbox and activate t

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_\text{samples}} \sum_{i=1}^{n_\text{samples}} \mathbb{1}[y_i = \hat{y}_i]$$

indicator funct. $= \begin{cases} 1 & \text{if } y_i = \hat{y}_i \\ 0 & \text{otherwise} \end{cases}$

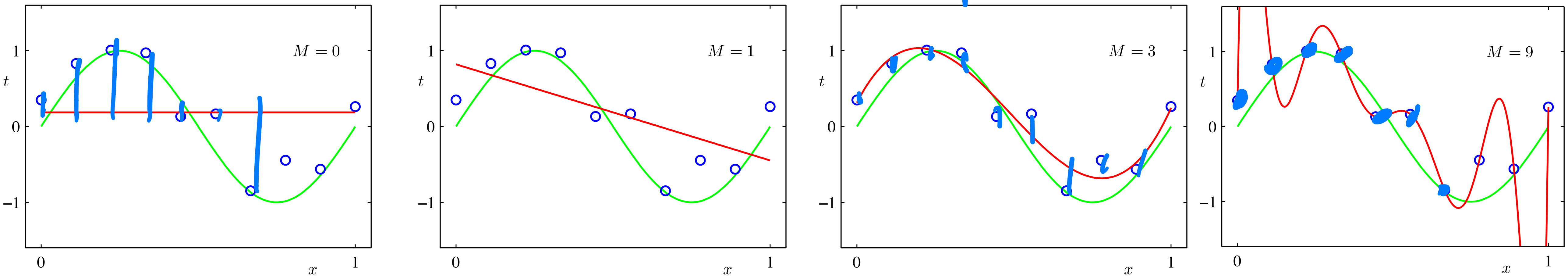
P: Performance measure

Regression

large MSE

small MSE

MSE = 0



mean squared error

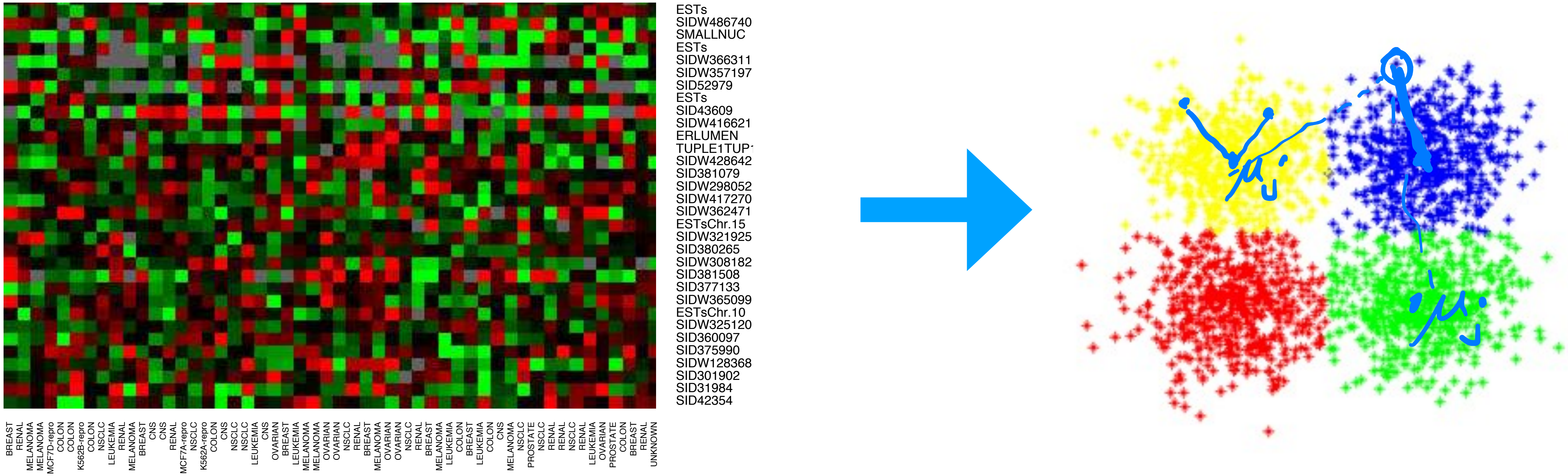
$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} (y_i - \hat{y}_i)^2$$

$\uparrow f_{\underline{w}}(x_i)$

Polynomials of order *M* (red) fit to data constructed as t = sin(2πx) + ε (green)

P: Performance measure

Clustering



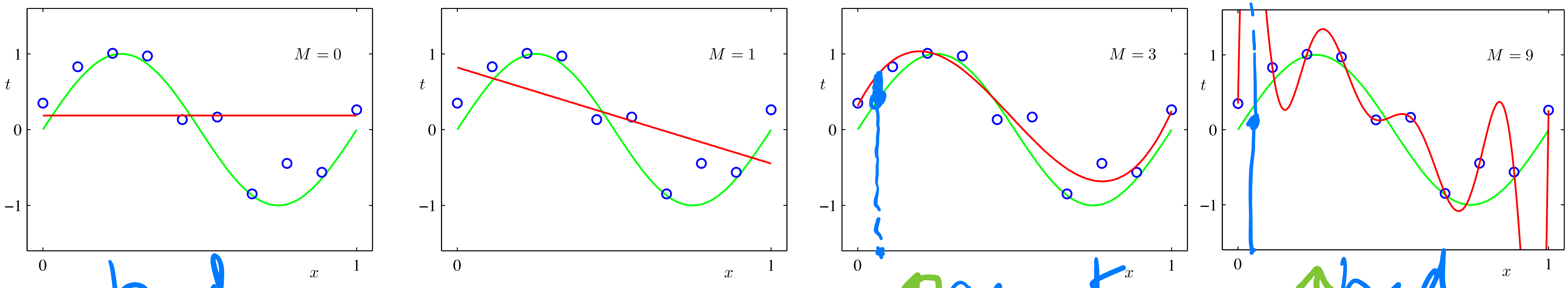
$$\text{within cluster sum of squares} = \sum_{i=1}^{n_{\text{samples}}} \underline{\min_{\mu_j \in C}} \, \|\mu_j - x_i\|^2$$

Expression matrix of genes (rows) for 64 human tumor samples (columns). [source: ESL 1.3]

P: Performance measure

overfitting

$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} (y_i - \hat{y}_i)^2$$



bad

great

bad

Best performance on training set :

Best performance on new datapoints :

Q: On which datapoints should performance be measured?

Generalisation:

performance should be measured on new data (test data)

What is machine learning?

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.”

- Tom M. Mitchell

Machine Learning, Tom Mitchell, McGraw Hill, 1997